# Projective Epipolar Rectification for a Linear Multi-imager Array

**CONFIDENTIAL: Submitted to 3DPVT 2010**


Zeyu Li
EECS, UC Berkeley
Berkeley, CA 94720
zeyuli@eecs.berkeley.edu

Harlyn Baker
Hewlett-Packard Laboratories
Palo Alto, CA 94304
harlyn.baker@hp.com

Gregorij Kurillo
EECS, UC Berkeley
Berkeley, CA 94720
gregorij@eecs.berkeley.edu



## Abstract

*Analysis of synchronized real-time multi-camera video can be facilitated by calibration tailored to its intended use. Calibration methods appropriate for two-view stereo processing become less so for tasks such as dense linear-array synchronous imaging for metric reconstruction or stereoscopic visualization where additional constraints apply. Calibrating for the construction of an Epipolar-Plane Image (EPI) volume, we introduce a novel method – based on a rank constraint on image homologies – for simultaneous high quality epipole estimation. Epipoles form the basis for recovering the fundamental matrices and, eventually, epipolar rectification parameters for the system. We present this new approach, discuss its characteristics in comparison with traditional metric multi-camera calibration, and show its utility in 3D display.*


## 1. Introduction

Acquisition, analysis, and presentation of three dimensional scene information arise in areas including environment capture, autonomous navigation, immersive interaction, 3D cinema, gaming, augmented reality, and remote collaboration. Multi-viewpoint imagery has a central role here in providing the three-dimensional world viewport. Synchronized real-time multi-camera capture to support this is now available, and issues in calibrating these systems present an immediate hurdle to the proper exploitation of the data they deliver. For many applications such as capturing and communicating video for 3D interaction the number of required cameras can grow quite large (tens to hundreds). With these cameras come more as well as different constraints. A number of calibration methods use planar targets for robustly recovering multi-camera parameters [1, 2, 3, 4]. We take a related approach with different objectives.

Our work deals with a multi-camera system specially configured (i.e., not generally positioned) for linear stereo with 3D display – and we want the best estimate of camera rectification parameters that serve this configuration. Not satisfied with pose estimates from traditional metric calibration [5] when used for epipolar alignment [6] (see Figure 4d), we suggest it's more important that ganged rectification be enabled than that derived pose parameters be respected. The problem lies partly in good estimation of the centers of projection and epipoles across pairs of cameras. Epipoles form the basis for recovering the fundamental matrices of a multi-imager camera system and for the rectification needed in both metric reconstruction and 3D display, so getting them right is important.

Figure 1. Simultaneous panoramic+multiview camera

An application of our camera system calls for multi-imager mosaicking for wide panoramic capture coupled with multi-viewpoint imaging for 3D acquisition. The device capturing these data is shown in Figure 1. The calibration requirements of these two cases are orthogonal: mosaicking analysis works in projective space (with homographies) while 3D calls for metric analysis; mosaicking tends to minimal feature overlap while multiview tends to maximal. Merging these, we wish to see if it is feasible to attain quality calibration in both spaces through upgrade of the simpler projective homographies where point-feature correspondence of metric analysis may be almost unavailable.

We introduce a novel method for high quality estimation of epipoles across a collection of imagers. For our special case – where the baseline of the system is known to be lin-

ear – optimization tunes these estimates to enable ganged analysis of the data with epipolar properties. This case has great value in two situations in particular: EPI-type reconstruction [7] and autostereoscopic 3D display [8]. While short of full metric character, the data transformations we present facilitate bulk multi-baseline ranging analysis and re-presentation of the imagery for free-viewpoint autostereo display. Our goal here is to show that we can facilitate these stereo and 3D display tasks through a coupled analysis involving simple but highly redundant use of image homographies. In this way, we can optimize our results for the measures that most concern us – epipolar alignment to facilitate ranging and 3D display – and bypass other pose parameters whose inclusion may interfere with deriving the transforms that best support this rectification.

Section 2 describes the role of calibration in the areas we are addressing and provides background on related methods for considering the problem of image calibration and epipolar alignment. In section 3 we present our calibration innovation for recovering a robust estimate of the fundamental matrix using a rank constraint on a derived image homology. Section 4 applies this development to epipolar rectification of a multi-imager camera array. In section 5 we highlight our experiments in evaluating this mapping. Section 6 uses this rectification for autostereo display. Finally, in section 7 we provide conclusions about this approach and sketch plans for further applying this analysis to passive multi-camera ranging.

## 2. Technical Background

The fundamental matrix ($F$) and epipoles ($e$ and $e'$) are key parameters relating stereo cameras. While a planar calibration target provides a simple mechanism for determining an homography relating camera and scene geometries, used singly it is inadequate for determining parameters such as the fundamental matrix and epipoles [1]. We ask a question here that has been asked before – can we use multiple instances of planar targets to jointly determine these parameters for a set of cameras. We answer this in a way that is different from others and present a novel constraint on the solution space. We use *homographies* to form objects called *homologies* that reveal inter-camera epipoles in a more direct and robust manner than previous efforts have delivered. These epipoles and associated homographies give us superior measures of the imaging system's fundamental matrices, and a special optimization on these permits us to rectify images for stereo and 3D display purposes. We provide a description and visualization of these geometric objects, and detail the mathematics around the homographies and homologies of the analysis.

### 2.1. Previous Work

Luong [1] present a method for estimating $F$ from several homographies. The calibration scene contains multiple planes, for each of which an $H$ can be robustly estimated using either the Direct Linear Transform or nonlinear methods. They then use the compatibility condition on $F$ and $H$: $F^T H$ should be an anti-symmetric matrix, so $F^T H + H^T F = 0$. Each homography matrix $H$ can provide 6 linear equations for $F$. Since $F$ has 8 degrees of freedom – up to scale – 2 or more planes ($H$'s) are needed to determine $F$. However, because $F$ is a rank-deficient matrix, it has sensitivities that make reliable estimation difficult. They suggest that increasing the number of planes beyond three is counter productive [1], a conclusion we challenge; by using separate images for plane observations, our solution can ensure constant quality homographies, so we continue to gain from redundancy.

Zelnik-Manor [9] exploit homologies in their 3D calibration work. Given two homographies $H_1$ and $H_2$ between two cameras induced by two planes, we point out that the *homology* $H_2^{-1} H_1$ can be parameterized as a "*rank-one perturbation of identity (ROPI) matrix*" (derived from a *rank-one perturbation of a rotation (ROPR) matrix* [8]). This matrix provides special information, as we now describe.

### 2.2. Homographies

An homography serves as a transfer function between projections of a 3D plane $\pi_1$ across a pair of cameras, say mapping points in image 1 to points in image 2 via the homography computed for $\pi_1$. If we have two such planes $\pi_1$ and $\pi_2$ (Figure 2), we can map points separately via their two derived homographies, $H_1$ and $H_2$. Mapping points from image 1 to image 2 through $H_1$ and then back to image 1 using the inverse of $H_2$, gives something resembling an identity – points return to the image from which they originated. But only some features return to their original positions, and these are of two types: (a) the image $v$ of the line of intersection of the two planes, v, and (b) the single point, $e$, located at the epipole in image 1 (defined by the relative positions and orientations of the two image planes). That these features are *fixed points* of the homology provides us a means to estimate them from a set of homologies.

The difficulty in using homographies (and homologies) is that they are defined only up to scale. Zelnik-Manor [9] referenced this point, and used the fact that each homology matrix has two degenerate unit eigenvalues to normalize them, removing the scale. However, in a real scenario with noise, it is problematic to assess such eigenvalue equality. Singular values, being on the order of the square root of the eigenvalues, provide a more stable environment for these calculations.

Our concern here is with estimation of $e$. While this ho-

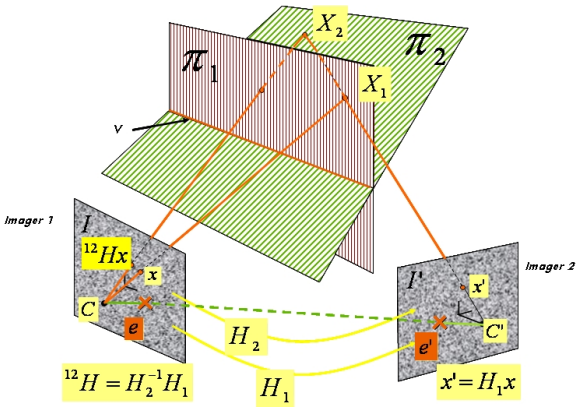


Figure 2. Imagers $I$ and $I'$, transfer planes $\pi_1$ and $\pi_2$, camera centers $C$ and $C'$, epipoles $e$ and $e'$, transfer point $x$ and $x'$ via plane points $X_1$ and $X_2$, and homographies $H_i$ with homologies ${}^{a,b}H$

mology fixed-point constraint provides a way to estimate epipoles with high precision, normalization issues make it difficult to use redundancy in these matrices to derive good estimates. Below, we show how a character of the homology matrix permits a more robust normalization, better epipoles, and more robust $F$s.

## 3. Fundamental Matrices

In this section, we discuss our methods for robustly estimating the fundamental matrices for a linear multi-camera array. We discuss the simplest case of two planes and two cameras, elevating to multiple planes and cameras, and the constraints that provide robust epipole and fundamental matrix estimation.

### 3.1. Homology matrix and normalization

Given two cameras $C$ and $C'$, with projection matrices $P = K[I, 0]$ and $P' = K'[\mathbf{R}, \mathbf{t}]$, respectively, the homographies induced by plane $\pi_1 = (n_1^T, d_1)^T$ and $\pi_2 = (n_2^T, d_2)^T$ can be parameterized as:

$$\begin{cases} H_1 \sim K'(\mathbf{R} - \mathbf{t}\frac{n_1^T}{d_1})K^{-1}, \\ H_2 \sim K'(\mathbf{R} - \mathbf{t}\frac{n_2^T}{d_2})K^{-1} \end{cases}$$

respectively. Where $(\mathbf{R}, \mathbf{t})$ is the Euclidean transformation of the camera $C'$ with respect to $C$, $d_1$ and $d_2$ are the perpendicular distances from the camera to the planes, and the equality $\sim$ means up to scale. The homology matrix, denoted as: ${}^{1,2}H = H_2^{-1}H_1$, can be decomposed [9, 10, 11] as:

$${}^{1,2}H = s(I - \lambda e v^T) \tag{1}$$

where $s$ is a scaling factor, $I$ is a 3 by 3 identity matrix and

$$\lambda \equiv \frac{1}{1 + \frac{n_2^T}{d_2}R^{-1}t}, \quad e \equiv KR^{-1}t, \quad v^T \equiv \left(\frac{n_1^T}{d_1} - \frac{n_2^T}{d_2}\right)K^{-1}$$

Zelnik-Manor [9] and Criminisi [10] pointed out that $e$ is the epipole up to scale, and $v$ is the image of the intersection of the two planes in camera $C$. Of course, the epipole $e$ depends only on the camera configuration, and the image $v$ of the intersection of $\pi_1$ and $\pi_2$, v, depends only on the scenes (i.e., the planes). This separability inspires us to consider the whole multi-camera array as a unified system. However, with use of homogeneous coordinates, there is an intrinsic scale factor $s$ in Eqn.(1) and, in order to put multiple-cameras and multiple-planes together, we have to normalize the homologies to an appropriate scale level.

Eqn.(1) defines a family of matrices of the form $A = I - xy^T$, which we term a *Rank-One Perturbation of Identity (ROPI)* matrix. It is easy to prove that the eigenvalues of a ROPI matrix have the form $(\mu, 1, 1)$ up to scale, i.e., it has two repeated and one non-repeated eigenvalues. Zelnik-Manor [9] and others [3] use this repeated eigenvalue to normalize the homology ${}^{1,2}H$, that is, $\mu = \frac{1}{\text{repeated\_eigenvalue}}$. In real applications with noise, however, it is numerically unstable to find this repeated eigenvalue, especially for the case where the three eigenvalues are similar (for example, for a larger-baseline camera rig).

To get a more robust normalization factor $s$, we establish the following property: *The ROPI matrix $I + xy^T$ has second singular value 1.* Therefore the second singular value $\sigma_2$ of ${}^{1,2}H$ can serve as a better normalizer. We obtain:

$${}^{1,2}\underline{H} = \frac{1}{\sigma_2}({}^{1,2}H) = I - \lambda e v^T \tag{2}$$

where ${}^{1,2}\underline{H}$ is the normalized homology. On eliminating $s$, we normalize all the homology matrices to a unified scale. This enables us in the next section to consider multiple-plane and multiple-camera elements in a single framework. Elsewhere, we have demonstrated methods for computationally stable solution of this singular value problem [8].

### 3.2. Generalizing the solution

With the above section showing normalization of the homology for the two-camera two-plane situation, here we generalize to the multiple-camera multiple-plane scenario. We have $C$ cameras observing $P$ planes. Without loss of generality, we choose one camera $C_0$ as our reference. The notations we use are: For a camera-pair $C_c$ and $C_0$, the homography induced by plane $\pi_p$ is ${}^pH_c$, the homology between plane $\pi_p$ and $\pi_r$ is ${}^{pr}H_c$, where $p, r = 1, \ldots, P$ and $c = 1, \ldots, C$. ${}^{pr}\underline{H}_c$ is the normalized version of the homology ${}^{pr}H_c$, and ${}^{pr}\tilde{\underline{H}}_c = {}^{pr}\underline{H}_c - I$.

Repeatedly considering Eqn.(2), we obtain:

$$\left\{ {}^{pr}\tilde{\underline{H}}_c = \lambda_{rc} e_c v_{pr}^T, \quad c = 0, 1, \ldots, C, \quad p, r = 1, 2, \ldots, P \right\}$$

where $\lambda_{rc} = \frac{1}{1+\frac{n_r^T}{d_r} R_c^T t_c}$, $e_c = K_0 R_c^T t_c$ is the epipole between camera $C_c$ and reference camera $C_0$, $v_{pr}^T = \left(\frac{n_p^T}{d_p} - \frac{n_r^T}{d_r}\right) K_0^{-1}$ is the image of the intersection of plane $\pi_p$ and $\pi_r$ in $C_0$, which will be shared by multiple cameras. We estimate all epipoles $e_c$ through this minimization:

$$\min_{\substack{||e_c||_2^2=1,\\ ||v_{pr}||_2^2=1}} J(\{\lambda_{rc}\},\{e_c\},\{v_{pr}\}) = \sum_{r=1}^{P}\sum_{p=1}^{P}\sum_{c=1}^{C} ||{}^{pr}\widetilde{\underline{H}}_c - \lambda_{rc} e_c v_{pr}^T||_F^2 \tag{3}$$

Unity constraints $||e_c||_2^2 = 1$ and $||v_{pr}||_2^2 = 1$ eliminated scale ambiguity in Eqn.(3). We seek $e$, not $v$, but solve for both because of their connection through the joint constraint. Also, it is clear that, for $p = r$, we have ${}^{pr}\underline{H} = ({}^{rp}\underline{H})^{-1}$ and ${}^{pr}\tilde{\underline{H}} = 0_{3\times 3}$.

In the next section, a general solution is given first and then several special cases are considered. After estimating epipoles, we can easily calculate the fundamental matrices.

### 3.3. Deriving the epipoles

To facilitate solving for the optimization in Eqn.(3), we define some simplifying notation.

$$\begin{cases} H_{rc} = \begin{pmatrix} {}^{1r}\widetilde{\underline{H}}_c & {}^{2r}\widetilde{\underline{H}}_c & \dots & {}^{Pr}\widetilde{\underline{H}}_c \end{pmatrix} \\ V_r = \begin{pmatrix} v_{1r}^T & v_{2r}^T & \dots & v_{Pr}^T \end{pmatrix}^T \end{cases} \tag{4}$$

The objective function in Eqn.(3) can be re-written:

$$\min_{\substack{||e_c||_2^2=1,\\ ||V_r||_2^2=1}} J(\{\lambda_{rc}\},\{e_c\},\{V_r\}) J = \sum_{r=1}^{P}\sum_{c=1}^{C} ||H_{rc} - \lambda_{rc} e_c V_r^T||_F^2 \tag{5}$$

Different methods can be used to solve this type of constrained optimization, for example gradient descent. Extending to a large number of cameras, we propose an alternate least-squares method with a rapid-converging monotonic-decreasing objective function.

**Theorem 1.** *The optimal value for $\lambda_{rc}$ is:*

$$\lambda_{rc}^* = V_r^T H_{rc}^T e_c \tag{6}$$

*Proof.* The objective function Eqn.(5) can be further written as:

$$J(\lambda_{rc}) = \sum_{r=1}^{P}\sum_{c=1}^{C} tr\left(H_{rc}^T H_{rc} - 2\lambda_{rc} V_r^T H_{rc}^T e_c + \lambda_{rc}^2\right)$$

Setting the first derivative about $\lambda_{rc}$ to zero, we have:

$$\frac{\partial J}{\partial \lambda_{rc}} = 2\lambda_{rc} - 2 \cdot V_r^T H_{rc}^T e_c \equiv 0$$

Therefore, $\lambda_{rc}^* = V_r^T H_{rc}^T e_c$. □

**Theorem 2.** *Let us define:*

$$A_r \equiv \sum_{c=1}^{C} H_{rc}^T e_c e_c^T H_{rc}, \quad B_c \equiv \sum_{r=1}^{P} H_{rc} V_r V_r^T H_{rc}^T \tag{7}$$

*then the optimal $V_r^*$ and $e_c^*$ are the singular vectors corresponding to the largest singular values of matrices $A_r$ and $B_c$, respectively.*

*Proof.* Substituting $\lambda_{rc}^*$ in Eqn.(6) into Eqn.(5), we can rewrite our objective function as:

$$\begin{aligned} J &= \sum_{r=1}^{P}\sum_{c=1}^{C} tr\left(H_{rc}^T H_{rc} - 2\lambda_{rc}^* H_{rc}^T e_c V_r^T + (\lambda_{rc}^*)^2\right) \\ &= \sum_{r=1}^{P}\sum_{c=1}^{C} tr\left(H_{rc}^T H_{rc} - V_r^T H_{rc}^T e_c V_r^T H_{rc}^T e_c\right) \end{aligned}$$

Thus, writing the minimization as a maximization problem:

$$\max_{\substack{||e_i||_2^2=1,\\ ||V_r||_2^2=1}} F(\{e_c\},\{V_r\}) = \sum_{r=1}^{P}\sum_{c=1}^{C} \left(V_r^T H_{rc}^T e_c V_r^T H_{rc}^T e_c\right) \tag{8}$$

which we further rewrite separately as:

$$\max_{||V_r||_2^2=1} F(\{V_r\}) = \sum_{r=1}^{P} V_r^T \left(\sum_{c=1}^{C} H_{rc}^T e_c e_c^T H_{rc}\right) V_r \tag{9}$$

$$\max_{||e_c||_2^2=1} F(\{e_c\}) = \sum_{c=1}^{C} e_c^T \left(\sum_{r=1}^{P} H_{rc} V_r V_r^T H_{rc}^T\right) e_c \tag{10}$$

Based on the definitions of $A_r$ and $B_c$, from the first maximization, the optimal $V_r^*$ is the eigenvector of $A_r$ corresponding to its largest eigenvalue, and the optimal $e_c^*$ is the corresponding largest-eigenvalue eigenvector of $B_c$. □

### 3.4. Best estimate of fundamental matrices

After obtaining the epipoles $e_c$ in Section 3.3, it is easy to estimate the $e_c'$ on camera $C_c$: $e_c' \sim ({}^pH_c)e_c$. The fundamental matrices $F_c$ can be calculated as $F_c = [e_c']_\times ({}^pH_c)$ or $F_c = ({}^pH_c)^{-T} \cdot [e_c]_\times$ [11].

To increase accuracy, we average $F_c$ for camera $C_c$ across all planes:

$$F_c^* = \frac{1}{P}\sum_{p=1}^{P} ({}^pH_c)^{-T} \cdot [e_c]_\times. \tag{11}$$

Because $[e_c]_\times$ is a rank-two asymmetric matrix and ${}^pH_c$ is a full rank matrix, $F_c^*$ will automatically be rank two, as required. Note that as a mathematical convenience in our formulation we constrain $e_c$ and $V_r$ to have unit length.

Although we minimize algebraic rather than geometric error, this works well in our application (see the evaluations in section 5). The algorithm runs very fast, with only $3 \sim 5$ iterations needed for convergence.

## 4. Linear array rectification

Recall that our goal is to have rectified EPI data for multiview stereo and 3D display. Here, we rectify our linear camera array using the fundamental matrices as estimated above. Our method consists of: 1) forming an initial estimate from F's; 2) refining the result using a non-linear optimization focussing around EPI alignment and linearity constraints. For comparison purposes, we also employ a full geometric calibration baseline algorithm as reference.

### 4.1. Initial mappers from $F$'s

Given a multi-camera array, $\{C_0, C_1, \ldots, C_C\}$, $C_0$ our reference camera, we seek a set of homographies $\hat{H}_i$, $i = 0, \ldots, C$, such that after applying $\hat{H}_i$ to $C_i$, the mapped images will have collinear scanlines where disparities are only horizontal.

For a post-rectified stereo pair, a fundamental matrix should have the form:

$$\tilde{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Therefore, we have:

$$\hat{H}_0^T \tilde{F} \hat{H}_i \sim F_i \tag{12}$$

There are infinitely many solutions to this [11, 12]. Mallon [12] estimates $\{H_i\}$ by projecting the epipole to infinity. However this only works in the two-camera situation (is questionable there) and is inadequate for our case. Here, we use the following strategy:

(1) We assume $\hat{H}_0$ to be an identity matrix, i.e., we leave the image of the reference camera unchanged. This minimizes image distortion due to the mappings and is a reasonable initial estimate in this linear-camera generally-orthogonally-viewing work (redressing this assumption can be accommodated later)

(2) For each camera $C_i$, we solve for its corresponding $\hat{H}_i$ using Eqn.(12):

$$\hat{H}_i = \begin{bmatrix} 1 & 0 & 0 \\ F_i^{(3,1)} & F_i^{(3,2)} & F_i^{(3,3)} \\ -F_i^{(2,1)} & -F_i^{(2,2)} & -F_i^{(2,3)} \end{bmatrix}$$

The first row is set to be $(1, 0, 0)$ [12] because there are no constraints on these components. We will further optimize over them in our next refinement stage. This formulation ensures $y$-direction error is small.

(3) The above solution ignores the horizontal shift along $x$. Because we are dealing with multiple cameras, we have to consider $x$ explicitly, which leads to the final initial estimate of the homography:

$$\hat{H}_i = \begin{bmatrix} 1 & 0 & d_i \\ F_i^{(3,1)} & F_i^{(3,2)} & F_i^{(3,3)} \\ -F_i^{(2,1)} & -F_i^{(2,2)} & -F_i^{(2,3)} \end{bmatrix}$$

Intuitively speaking, $d_i$ controls the horizontal shift of each image, which actually depends on both the focal length of the camera and the scene depth. Additionally, $d_i$ is not unique: If $d_i$ is a solution, then $2 \times d_i$ is also a solution. To better accommodate our scene depth, $d_i$ is initialized to the average distance along the $x$ direction, given multiple corresponding points. Again, these parameters are refined during the optimization.

Having established initial estimates for our homographies, we further parameterize them as follows:

$$\hat{H}_i = \begin{bmatrix} 1 & h_i^{(1)} & h_i^{(2)} \\ h_i^{(3)} & h_i^{(4)} & h_i^{(5)} \\ h_i^{(6)} & h_i^{(7)} & h_i^{(8)} \end{bmatrix}$$

where the (1,1)-component has been fixed to 1 to correct the homography's scale ambiguity. Next, we optimize this set of parameters through a non-linear refinement.

### 4.2. Non-linear refinement process

Epipolar rectification of a multi-camera array brings two objectives to mind: 1) vertical disparity should be zero (image displacement being strictly horizontal); 2) the EPI linearity property means points corresponding across cameras will be linear in rectified EPI-volume space.

Our objective function is based on these two errors:

$$J(\theta) = J_{vert} + \lambda J_{linear} \tag{13}$$

where $\lambda$ is a pre-defined trade-off parameter. First, we define the vertical error, then the linear.

With $N$ corresponding points across all $C$ cameras, $\{x_{c,i}, y_{c,i}\}$, where $c = 1, \ldots, C$ and $i = 1, \ldots, N$, applying $\hat{H}_c$ to cameras $C_c$ gives transformed image points $i$:

$$\begin{bmatrix} x'_{c,i} \\ y'_{c,i} \\ 1 \end{bmatrix} = \frac{\hat{H}_c \times [x_{c,i}, y_{c,i}, 1]^T}{[0, 0, 1] \times \hat{H}_c \times [x_{c,i}, y_{c,i}, 1]^T}$$

where the RHS denominator arises from use of homogeneous coordinates. Therefore, to horizontally align the corresponding transformed image points we minimize:

$$J_{vert} = \sum_{i=1}^{N} \sum_{c=1}^{C} \left(y'_{c,i} - \bar{y'_i}\right)^2, \quad \text{where } \bar{y'_i} = \frac{1}{C} \sum_{c=1}^{C} y_{c,i} \tag{14}$$

Corresponding points incur minimal vertical disparity.

For the linearity error, we consider that EPI-slice corresponding points [7] should form a line whose slope is determined by the baseline and depth. We denote the baseline of camera $C_c$ with respect to reference camera $C_0$ as $b_c$, then given a set of corresponding points, the following points in x-z space should be a line:

$$\{p_{c,i} \equiv \{x'_{c,i}, b_c\}\}, \quad c = 1, \ldots, C, \quad \forall i$$

Fitting line $l_i$ to these points:

$$J_{linear} = \sum_{i=1}^{N} \sum_{c=1}^{C} [\text{dist}(p_{c,i}, l_i)]^2 \tag{15}$$

with $\text{dist}(p, l)$ the orthogonal distance between $p$ and $l$.

For the calibrated situation, we can explicitly include estimated baseline values. However, in the uncalibrated scenario here, without pose information we invoke an equidistant baseline assumption which, given our means of construction, is sufficiently accurate (to fractions of a millimeter). As usual, with point sets $x$, normalization drives their mean to zero and average distance to $\sqrt{2}$. This prevents the larger baseline-displacement component from dominating the optimization.

### 4.3. Baseline algorithm

Our baseline algorithm starts with full geometric checkerboard calibration [5]. It differs from the standard method in how it initializes the camera homographies.

(1) Fit a line $l$ to the camera centers minimizing the orthogonal distance to all camera centers. $l$, represented by a 3-D point $A$ and direction $a$, is chosen as the baseline (Figure 3a).

(2) Rotate $C_0$ so that its $x$-axis is aligned with $a$. Minimizing the distortion due to rotating images, we orthogonally project the $z$-axis of the reference camera $C_0$ onto plane $P$, and treat this as our new $z'$ axis. The $y'$ axis is the cross-product of $x'$ and $z'$. Denoting $R_w = [x'; y'; z']$, we move our world coordinate frame to $W$ (Figure 3b).

(3) Calculate the pose of each camera with respect to $W$, rotating each camera so its optical axis is perpendicular to $x'$ and parallel to $z'$ (Figure 3c). This rotation process can be implemented as an homography.

Essentially, we approximate an ideal *multi-baseline system* by the cameras in Figure 3c without moving their centers of projection. How well this works depends on the distance each camera moves in $z'$. Since $l$ is the direction minimizing this distance for all camera centers, the effect of moving camera centers has been minimized.

The initial estimate of $\{\hat{H}_i\}$ is refined by solving Eqn.(13) in section 4.2.

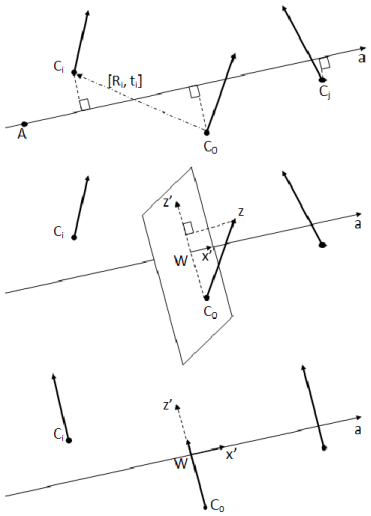


Figure 3. Multi-camera rectification for general scene: a) initial estimates of $\{H_i\}$, b) orientations and c) alignments.

## 5. Evaluation

Section 4.2 produced a set of transforms derived from homographies and adjusted by a non-linear refinement exploiting our known configuration constraints. Section 4.3 produced a set of transforms derived from metric calibration and then refined through a construction argument for epipolar rectification. We have conducted two sets of experiments in evaluating the quality of these alignment-based rectifications. We assess – for a 10-imager multi-camera system – the accuracy of estimating fundamental matrices in the projective case (4.2), the reprojection error for the metric case (4.3 initialization), and the overall rectification errors for both cases. The latter includes applying a post-adjustment non-linear refinement to the metric case, seeing how the vertical disparity and epipolar linearity constraints of 4.2 might help enhance the metric results of 4.3.

Rectification error measurement employs the corner points of imaged checkerboards. We selected the corner method since it provides for metric comparison; doing a single projective calibration where there may be no point correspondences and seeing how it might be upgraded through other constraints has been the motivation for this study.

### 5.1. Accuracy of fundamental matrices

The fundamental matrix maps points between images. The metric we are using to evaluate error is the following: given known corresponding points in the two images, $x$ and $x'$, we calculate the orthogonal distance of $x'$ deviating from the appropriate $F$-estimated epipolar line, $Fx$. This gives us an average RMS for all cameras of $0.1874$ pixel. Figure 4a shows the F-error for each individual camera. In addi-

tion to showing how effective the non-linear optimization is, these plots indicate that the non-metric method gives us very good fundamental matrix estimations, though we do not explicitly minimize this error. This is accounted for by the strong global constraints among all cameras, where our robust $e$'s deliver $F$'s whose accuracy-maximizing norm we can select.

To test the stability of our algorithm, we ran the following experiments: for each feature point we add Gaussian noise: $n \sim N(0, \sigma)$, selecting $\sigma$ in $\{0.05 : 0.05 : 0.7\}$, and perform 30 randomized runs. Figure 4b reports the average error and its standard deviation. That the estimation error is linear with respect to input error shows the F-estimation algorithm to be robust to noise.

### 5.2. Geometric calibration

We use a variation of the standard checkerboard method to calibrate our multi-camera array, and examine *reprojection error* to assess its apparent quality. The mean error for the ten-camera system employed is $0.2685$ pixel, with standard deviation $0.198$ pixel and a max error of $1.6576$ pixels (Figure 4c). Recovered camera layout is shown in Figure 4d. This shows graphically that the quality of geometric calibration need not be what the numbers indicate: for a system that we know to be aligned within millimeters, camera centers are estimated to have about a half inch of vertical deviation. Part of the reason for this variance is that geometric calibration involves a non-linear optimization where pose and intrinsics are intertwined. One goal of our developments here has been to show that, solving more simply for epipoles, we are able to do better on our major error concern – rectification alignment.

### 5.3. Accuracy of rectification

Using a metric similar to that of section 6.2, we also compare the final error, in pixels, along the $y$ direction (see Table 1). It can be seen that our projective method gives better both initial and final rectification results. Using the same non-linear optimization, our considerably simpler homography-based solution shows the value of exploiting this ROPI-driven least-squares epipole-estimation solution.

Figure 5 shows a set of selected epipolar lines across a band in 10 rectified images. Good results are observed where there is no noticeable vertical deviation in the selected features – scan this figure to verify the alignment quality. Figure 6a shows a single image from this set, indicating the features selected to demonstrate alignment.

To check the linearity EPI property, we plot one EPI-slice (Figure 7). Clearly, we see lines, with slope determined by the depth of the corresponding points.

In addition to these metric evaluations, every use of the system described in section 6 presents an opportunity for users to assess the full pipeline quality; and it works.

Table 1. Rectification (y-direction) reprojection error

| **algorithm** | **initial error** | **final error** |
|---|---|---|
| baseline metric method | 0.778259 | 0.120604 |
| new projective method | 0.396695 | 0.118561 |

## 6. Autostereoscopic Display

We leave use of this method in stereo reconstruction for another time, and here briefly discuss the more experiential use of the system in autostereo display, sketching the video pipeline and the display implementation.

### 6.1. Rectification in 3D capture and display

Image rectification – reorienting and resampling so that epipolar lines become corresponding scan lines – is an essential step in structuring imagery for autostereo viewing. This mapping must correct for imprecision in component mechanical assembly, correct lens distortions, accommodate to the perspective keystoning that comes from having the capture cameras converge on the desired content, and minimize corresponding point vertical disparity (epipolar alignment). Calibrating the display side is easier, it just has to maintain the alignment delivered by the camera system. In our work, display calibration simply involves determining homographies.

### 6.2. Display Methods

Numerous approaches to autostereo display are chronicled in several conference series [13, 14]. Our approach [8] is a projection method based on a retroreflective material bonded to a diffusing layer that permits a viewer's eye located within a reflected diffusion zone before the screen to see the output of a single overhead-mounted projector (Figure 6b). Multiple projectors provide multiple adjacent view zones. This real-time autostereo facility is in operation with 18 cameras and 18 projectors delivering 16 discrete binocular view zones. It acts in communication with a 9-camera site [8], and as a mirror. Video bandwidth here approaches 2.5 gigabits per second. Our solution provides a unified single-PC approach, where cameras, displays and computation could be delivered for a bill-of-materials cost of under ten thousand dollars. The projective epipolar rectification we describe here is key to its quality and its simplicity.

## 7. Conclusions and Future Work

We have presented a novel algorithm to recover epipolar geometry for a multi-camera array free of metric requirements, and highlighted its use in autostereoscopic display. We are now expanding this capability to autostereo display of panoramic imagery, focussing on issues of perceptual

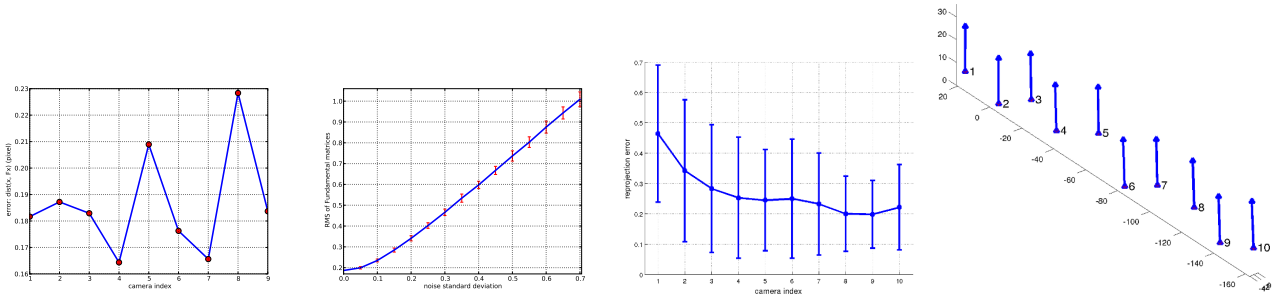


Figure 4. a) Fundamental matrix error in pixels along y-direction; b) Stability of estimation of F's; c) Reprojection error for each camera; d) Camera layout recovered from the geometric calibration (arrows represent optical axes).

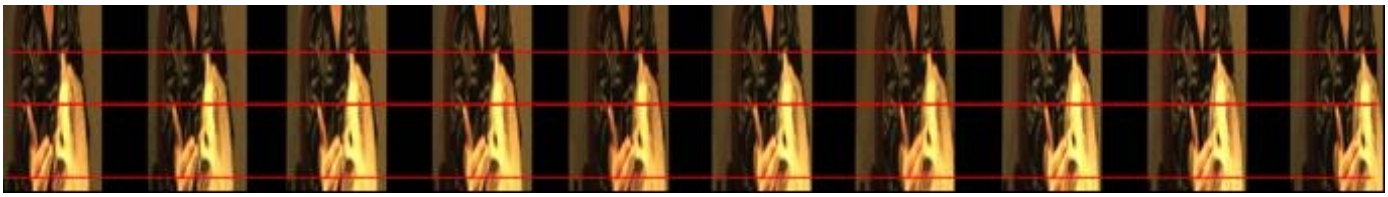

Figure 5. Rectification Result – epipolar lines across 10 imagers through the three selected features

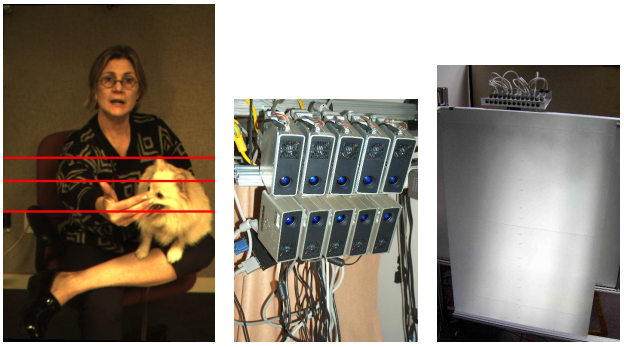

Figure 6. a) Reference image for rectification results: three selected features; b) Autostereo display: overhead projectors and diffusing retroreflection screen (notice cameras above screen)

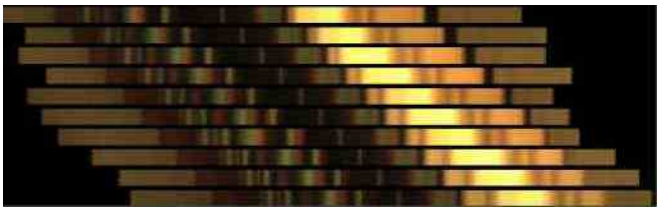

Figure 7. Rectification verification: EPI-slice near middle line of Figure 5 (boundaries indicate original horizontal alignment)

quality in an immersive experience. Our calibration studies will continue, as we work to develop hybrid analyses that couple 3D display with metric use in range recovery to support advanced user interaction activities.

## References


[1] Q.-T. Luong, O. Faugeras, Determining the fundamental matrix with planes: unstability and new algorithms, CVPR, 1993. 1, 2

[2] E. Malis, R. Cipolla, Camera self-calibration from unknown planar structures enforcingthe multiview constraints between collineations, PAMI 24:9, 2002. 1

[3] T. Ueshiba, F. Tomita, Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices, ICCV, 2003. 1, 3

[4] P. Chen, D. Suter, Rank constraints for homographies over two views: revisiting the rank four constraint, IJCV 81:2, 2009. 1

[5] Z. Zhang, Flexible Camera calibration by viewing a plane from unknown orientations, ICCV, 1999. 1, 6

[6] A. Fusiello, E. Trucco, A. Verri, A compact algorithm for rectification of stereo pairs, Machine Vision and Applications 12:1, 2000. 1

[7] R. Bolles, H. Baker, D. Marimont, Epipolar-plane image analysis, an approach to determining structure from motion, IJCV 1:1, 1987. 2, 6

[8] withheld for anonymity 2, 3, 7

[9] L. Zelnik-Manor, M. Irani, Multiview constraints on homographies, PAMI 24:1, 2002. 2, 3

[10] A. Criminisi, I. Reid, A. Zisserman, Duality, rigidity and planar parallax, ECCV, 1998. 3

[11] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, Cambridge Press, 2000. 3, 4, 5

[12] J. Mallon, P. Whelan, Projective rectification from the fundamental matrix, Image and Vision Computing 23:7, 2005. 5

[13] http://www.stereoscopic.org/ 7

[14] http://sp.cs.tut.fi/3dtv-con2010/ 7